

# ON THE UNIQUENESS OF SPARSE TIME-FREQUENCY REPRESENTATION OF MULTISCALE DATA\*

CHUNGUANG LIU<sup>†</sup>, ZUOQIANG SHI<sup>‡</sup>, AND THOMAS Y. HOU<sup>§</sup>

**Abstract.** In this paper, we analyze the uniqueness of the sparse time-frequency decomposition and investigate the efficiency of the nonlinear matching pursuit method. Under the assumption of scale separation, we show that the sparse time-frequency decomposition is unique up to an error that is determined by the scale separation property of the signal. We further show that the unique decomposition can be obtained approximately by the sparse time-frequency decomposition using nonlinear matching pursuit.

**Key words.** sparse time-frequency decomposition, scale separation, nonlinear matching pursuit

**AMS subject classifications.** 94A12, 42A20

**DOI.** 10.1137/141002098

**1. Introduction.** Today, data play a more and more important role in our lives. It has become increasingly necessary to develop effective data analysis tools to extract useful information from massive amounts of data. Frequency is one of the most important features for oscillatory data. In many physical problems, frequencies encode important information of the underlying physical mechanism. Many time-frequency analysis methods have been developed to extract information of frequencies and the corresponding amplitudes from the measurement of signals. These include the windowed Fourier transform, the wavelet transform [5, 15], the Wigner–Ville distribution [9], and the empirical mode decomposition (EMD) method [14, 17]. Among these different time-frequency analysis methods, the EMD method provides an efficient adaptive method to extract frequency information from nonlinear and nonstationary data, and it has been successfully used in many applications. However, due to its empirical nature, the EMD method still lacks a rigorous mathematical foundation. Recently, a number of attempts have been made to provide a mathematical foundation for this method; see, e.g., the synchrosqueezed wavelet transform [6], the empirical wavelet transform [8], and the variational mode decomposition [10].

In the last few years, inspired by the EMD method and compressive sensing [1, 2, 7], Hou and Shi proposed a novel time-frequency analysis method based on the sparsest representation of multiscale data [11]. In this method, the signal is decomposed into a finite number of intrinsic mode functions with a small residual:

$$(1.1) \quad f(t) = \sum_{j=1}^M a_j(t) \cos \theta_j(t) + r(t), \quad t \in \mathbb{R},$$

\*Received by the editors December 30, 2014; accepted for publication (in revised form) May 6, 2015; published electronically July 8, 2015. This work was supported by NSF FRG grants DMS-1159138 and DMS-1318377, AFOSR MURI grant FA9550-09-1-0613, and DOE grant DE-FG02-06ER25727.

<http://www.siam.org/journals/mms/13-3/100209.html>

<sup>†</sup>Department of Mathematics, Jinan University, Guangzhou, 510632, China (tcgliu@jnu.edu.cn). The research of this author was supported by NSFC grant 11371173.

<sup>‡</sup>Mathematical Sciences Center, Tsinghua University, Beijing, 100084, China (zqshi@math.tsinghua.edu.cn). The research of this author was supported in part by NSFC grant 11201257.

<sup>§</sup>Applied and Computational Mathematics, Caltech, Pasadena, CA 91125 (hou@cms.caltech.edu).

where  $a_j(t)$ ,  $\theta'_j(t) > 0$ ,  $j = 1, \dots, M$ , and  $r(t)$  is a small residual. We assume that  $a_j(t)$  and  $\theta'_j$  are less oscillatory than  $\cos \theta_j(t)$ . What we mean by “less oscillatory” will be made precise below by the definition of the scale separation property of the signal. Borrowing the terminology from the EMD method, we call  $a_j(t) \cos \theta_j(t)$  the intrinsic mode function (IMF) [14]. After the decomposition is obtained, the instantaneous frequencies  $\omega_j(t)$  are defined as

$$(1.2) \quad \omega_j(t) = \theta'_j(t),$$

and the amplitude is given by  $a_j(t)$ .

One main difficulty in computing the decomposition (1.1) is that the decomposition is not unique. To pick up the “best” decomposition among all feasible ones, Hou and Shi proposed to decompose the signal by looking for the sparsest decomposition by solving the following nonlinear optimization problem:

$$(1.3) \quad \begin{array}{ll} \text{Minimize} & M \\ (a_k)_{1 \leq k \leq M}, (\theta_k)_{1 \leq k \leq M} & \\ \text{subject to} & \left\| f - \sum_{k=1}^M a_k \cos \theta_k \right\|_{L^2} \leq \epsilon, \quad a_k \cos \theta_k \in \mathcal{D}, \end{array}$$

where  $\epsilon$  is the noise level and  $\mathcal{D}$  is the dictionary consisting of all IMFs (see [11] for its precise definition).

The idea of looking for the sparsest representation over the time-frequency dictionary has been exploited extensively in the signal processing community; see, e.g., [16, 3]. Comparing with other existing methods, the novelty of the method proposed by Hou and Shi is that the time-frequency dictionary being used is much larger. This method has the advantage of being fully adaptive to the signal.

The optimization problem (1.3) is nonlinear and nonconvex. It is challenging to solve this nonlinear optimization problem efficiently. To overcome this difficulty, Hou and Shi proposed an efficient algorithm based on nonlinear matching pursuit and fast Fourier transform to solve the above nonlinear optimization problem. In a subsequent paper [13], the authors proved the convergence of their nonlinear matching pursuit algorithm for periodic data that satisfy a certain scale separation property. Further, they have demonstrated the effectiveness of this method by decomposing realistic signals arising from various applications. However, from the theoretical point of view, one important question remains open: the uniqueness of the solution of the optimization problem (1.3). This is precisely the main focus of the current paper.

In this paper, we will show that under the assumption of scale separation, the solution of optimization problem (1.3) is unique up to an error determined by the scale separation property. First, we give a precise definition of scale separation of a signal as follows.

**DEFINITION 1.1** (scale separation). *One function  $f(t) = a(t) \cos \theta(t)$  is said to satisfy a scale separation property with a separation factor  $\epsilon > 0$  if  $a(t)$  and  $\theta(t)$  satisfy the following conditions:*

$$\begin{aligned} a(t) &\in C^1(\mathbb{R}), \quad \theta \in C^2(\mathbb{R}), \quad \inf_{t \in \mathbb{R}} \theta'(t) > 0, \\ \frac{\sup_{t \in \mathbb{R}} \theta'(t)}{\inf_{t \in \mathbb{R}} \theta'(t)} &= M' < +\infty, \quad \left| \frac{a'(t)}{\theta'(t)} \right| \leq \epsilon, \quad \left| \frac{\theta''(t)}{(\theta'(t))^2} \right| \leq \epsilon \quad \forall t \in \mathbb{R}. \end{aligned}$$

In the above definition, the first three assumptions are on the regularity of the envelope  $a(t)$  and the instantaneous frequency  $\theta'(t)$ . This regularity is relatively mild

and can be relaxed to a piecewise smooth function. The key assumptions on the scale separation property are the last two assumptions. The assumption  $|\frac{a'(t)}{\theta'(t)}| \leq \epsilon$  quantifies what we mean by the envelope being less oscillatory than the normalized signal  $\cos(\theta(t))$ . The last assumption,  $|\frac{\theta''(t)}{(\theta'(t))^2}| \leq \epsilon$ , essentially says that the frequency oscillation is relatively weak compared with the square of the frequency itself. In the previous theoretical study of time-frequency analysis, a much stronger assumption on the frequency oscillation is made, i.e.,  $|\frac{\theta''}{\theta'}| \leq \epsilon$ ; see, e.g., [6, 4]. In the problems that we consider, the instantaneous frequency  $\theta'(t)$  is typically quite large. Therefore, the assumption  $|\frac{\theta''(t)}{(\theta'(t))^2}| \leq \epsilon$  is much weaker than the typical assumption,  $|\frac{\theta''}{\theta'}| \leq \epsilon$ .

With the definition of scale separation, we construct the dictionary  $\mathcal{D}_\epsilon$  by putting all the functions satisfying scale separation together:

$$(1.4) \quad \mathcal{D}_\epsilon := \{a(\cdot) \cos \theta(\cdot) : (a, \theta) \in U_\epsilon\},$$

where

$$(1.5) \quad U_\epsilon := \left\{ (a, \theta) : a > 0, \theta' > 0; \left| \frac{a'}{\theta'} \right| \leq \epsilon, \left| \frac{\theta''}{[\theta']^2} \right| \leq \epsilon, \frac{\sup_{t \in \mathbb{R}} \theta'(t)}{\inf_{t \in \mathbb{R}} \theta'(t)} = M' < +\infty \right\}.$$

$\mathcal{D}_\epsilon$  is the dictionary that we will use to represent the signal. For any given signal  $f(t)$ , we decompose  $f$  over the dictionary  $\mathcal{D}_\epsilon$  by looking for the sparsest representation:

$$(P_0) \quad \begin{array}{ll} \text{Minimize} & M \\ \text{subject to} & \left| f(t) - \sum_{k=1}^M a_k(t) \cos \theta_k(t) \right| \leq \epsilon_0, \\ & a_k(t) \cos \theta_k(t) \in \mathcal{D}_\epsilon, \quad k = 1, 2, \dots, M. \end{array}$$

Here  $\epsilon_0$  is a given threshold of the accuracy of the decomposition. Typically,  $\epsilon_0$  is set according to the amplitude of noise.

To get the uniqueness, we need to assume that the signal  $f(t)$  is well separated, which is defined in Definition 1.2.

**DEFINITION 1.2** (well-separated signal). *A signal  $f : \mathbb{R} \rightarrow \mathbb{R}$  is said to be well separated with separation factor  $\epsilon$  and frequency ratio  $d$  if it can be written as*

$$(1.6) \quad f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t) + r(t),$$

where all  $f_k(t) = a_k(t) \cos \theta_k(t)$  satisfy the scale separation property with separation factor  $\epsilon$ ,  $r(t) = O(\epsilon_0)$ , and their phase functions  $\theta_k$  satisfy

$$(1.7) \quad \theta'_k(t) \geq d\theta'_{k-1}(t) \quad \forall t \in \mathbb{R},$$

and  $d > 1$ ,  $d - 1 = O(1)$ .

In the above definition, the key measurement of separation among different components is the value  $d$ , which measures the frequency gap among different components. For a well-separated signal, we can prove that the solution of the optimization problem  $(P_0)$  is unique up to an error  $O(\epsilon)$ . The main result is summarized in Theorem 1.3.

**THEOREM 1.3.** *Let  $f(t)$  be well separated with separation factor  $\epsilon \ll 1$  and frequency ratio  $d$  as defined in Definition 1.2. Then  $(a_k, \theta_k)_{1 \leq k \leq M}$  is an optimal*

solution of the optimization problem  $(P_0)$  and it is unique up to the error  $\epsilon$ ; i.e., if  $(\tilde{a}_k, \tilde{\theta}_k)_{1 \leq k \leq \tilde{M}}$  is another optimal solution of  $(P_0)$ , then  $\tilde{M} = M$  and

$$(1.8) \quad |a_k(t) - \tilde{a}_k(t)| = O(\epsilon), \quad \frac{|\theta_k(t) - \tilde{\theta}_k(t)|}{\theta'_k(t)} = O(\epsilon) \quad \forall t, \quad k = 1, \dots, M.$$

This theorem is proved by carefully studying the wavelet transform of each IMF. The details of the proof can be found in section 2.

We remark that there has been some very nice progress in developing a mathematical framework for an EMD-like method using synchrosqueezed wavelet transforms [6] and windowed Fourier transform [4]. These methods are based on continuous wavelet transform and windowed Fourier transform, respectively, and do not look for the sparsest decomposition directly. So the question of uniqueness is not the same as that we consider in this paper.

The rest of the paper is organized as follows. The uniqueness of the sparse time-frequency decomposition is analyzed in section 2. In section 3, we analyze the performance of an algorithm based on matching pursuit. Some concluding remarks are made in section 4. We defer a few technical lemmas to the appendices.

**2. Uniqueness of  $(P_0)$  for well-separated signals.** In this section, we assume that the signal,  $f(t)$ , satisfies the scale separation property with separation factor  $\epsilon$  and frequency ratio  $d$  as defined in Definition 1.2.

For this kind of signal, the existence of the solution of  $(P_0)$  is obvious. Since  $f(t)$  already has a representation,

$$(2.1) \quad f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t) + r(t),$$

this gives a feasible decomposition. Each feasible decomposition gives a positive integer; by collecting all these positive integers together, we get a set  $A$ . Then we know that  $A$  is nonempty and has a lower bound. Let  $M_0 = \inf A$ . Since  $A$  consists of positive integers,  $M_0 = \inf A$  can be achieved. Then the existence of the solution of  $(P_0)$  is proved.

But the uniqueness is much more complicated. In the next subsection, we will prove that if  $f$  is a well-separated signal with separation factor  $\epsilon$  and frequency ratio  $d > 1$  and  $\epsilon \ll 1$ , then the solution of  $(P_0)$  is unique up to  $\epsilon$  and  $\epsilon_0$ .

To simplify the notation, in the rest of this paper, we assume that  $\epsilon_0$  and  $\epsilon$  have the same order and denote both of them by  $\epsilon$ .

We prove Theorem 1.3 by carefully studying the wavelet transform of each IMF. As we show in Lemma 2.1, the continuous wavelet transform (CWT) of each IMF is confined in a narrow band by properly choosing an appropriate wavelet function. Then the uniqueness can be obtained by comparing the CWTs of different decompositions.

In order to complete the proof, first we need to estimate the width of the CWT of each IMF. This estimation is given in the following lemma.

LEMMA 2.1. *Let  $\psi$  be a wavelet function such that*

$$I_1 = \int_{\mathbb{R}} |\psi(\tau)| d\tau < +\infty, \quad I_2 = \int_{\mathbb{R}} |\tau \psi'(\tau)| d\tau < +\infty, \quad I_3 = \int_{\mathbb{R}} |\tau^2 \psi''(\tau)| d\tau < +\infty.$$

Suppose  $(a, \theta) \in U_\epsilon$ , i.e.,

$$\left| \frac{a'}{\theta'} \right| \leq \epsilon, \quad \left| \frac{\theta''}{[\theta']^2} \right| \leq \epsilon, \quad \frac{\sup_{t \in \mathbb{R}} \theta'(t)}{\inf_{t \in \mathbb{R}} \theta'(t)} = M' < +\infty;$$

then we have

$$\mathcal{W}(ae^{-i\theta})(t, \omega) = \frac{1}{\sqrt{\omega}} \int_{\mathbb{R}} a(\tau) e^{-i\theta(\tau)} \psi\left(\frac{\tau-t}{\omega}\right) d\tau = \sqrt{\omega} a(t) e^{-i\theta(t)} \widehat{\psi}(\omega\theta'(t)) + C\sqrt{\omega}\epsilon,$$

where  $C = (A + 4|a(t)| + 1)I_1 + [M' + (M' + 1)|a(t)|]I_2 + M'|a(t)|I_3$  and  $A = \sup_{t \in \mathbb{R}} |a(t)|$ .

The proof of this lemma can be found in Appendix A.

*Remark 2.1.* One sufficient condition to make sure that  $I_1, I_2, I_3 < +\infty$  is that the Fourier transform of the wavelet,  $\widehat{\psi} \in C^4(\mathbb{R})$ , has compact support. In the proof of Theorem 1.3, we use the wavelet whose Fourier transform is a fifth order B-spline function such that the above sufficient condition is satisfied.

**2.1. Proof of Theorem 1.3.** Now, we are ready to give the proof of Theorem 1.3.

*Proof of Theorem 1.3.* The details of the proof may be a bit tedious, but the idea is very clear. First, we assume there are two decompositions:

$$(2.2) \quad f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t) + O(\epsilon) = \sum_{k=1}^{\tilde{M}} \tilde{a}_k(t) \cos \tilde{\theta}_k(t) + O(\epsilon).$$

Using Lemma 2.1, we have

$$(2.3) \quad \begin{aligned} \omega^{-1/2} \mathcal{W}(f)(t, \omega) &= \sum_{k=1}^M a_k(t) e^{-i\theta_k(t)} \widehat{\psi}(\omega\theta'_k(t)) + O(\epsilon) \\ &= \sum_{k=1}^{\tilde{M}} \tilde{a}_k(t) e^{-i\tilde{\theta}_k(t)} \widehat{\psi}(\omega\tilde{\theta}'_k(t)) + O(\epsilon). \end{aligned}$$

By analyzing the support of  $\mathcal{W}(f)(t, \omega)$ , we can get  $M = \tilde{M}$  and

$$(2.4) \quad a_k(t) e^{-i\theta_k(t)} \widehat{\psi}(\omega\theta'_k(t)) = \tilde{a}_k(t) e^{-i\tilde{\theta}_k(t)} \widehat{\psi}(\omega\tilde{\theta}'_k(t)) + O(\epsilon), \quad k = 1, \dots, M.$$

Using these equalities, we can show that (1.8) is true.

First, we pick a specific wavelet function  $\psi$ . We require that its Fourier transform  $\widehat{\psi} \in C^4$  has support in  $[1 - \Delta, 1 + \Delta]$  where  $0 < \Delta < \frac{\sqrt{d}-1}{\sqrt{d}+1}$  and  $\widehat{\psi}(1) = 1$  is the maximum of  $|\widehat{\psi}|$ . In this proof, we choose  $\widehat{\psi}$  to be a fifth order B-spline function after proper scaling and translation.

For any  $\theta \in C^1$ , define

$$U_\theta = \left\{ (t, \omega) \in \mathbb{R}^2 : |\widehat{\psi}(\omega\theta'(t))| > \epsilon \right\}, \quad U_\theta(t) = \left\{ \omega \in \mathbb{R} : |\widehat{\psi}(\omega\theta'(t))| > \epsilon \right\}.$$

Now, fix a time  $t = t_0 \in [0, 1]$ , and for any  $l \in \{1, \dots, M\}$ , let  $\omega_{l,0} = \frac{1}{\theta'_l(t_0)}$ . Using (2.3), we have

$$(2.5) \quad \begin{aligned} \omega^{-1/2} \mathcal{W}(f)(t_0, \omega_{l,0}) &= \sum_{k=1}^M a_k(t_0) e^{-i\theta_k(t_0)} \widehat{\psi}(\theta'_k(t_0)/\theta'_l(t_0)) + O(\epsilon) \\ &= \sum_{k=1}^{\tilde{M}} \tilde{a}_k(t_0) e^{-i\tilde{\theta}_k(t_0)} \widehat{\psi}(\tilde{\theta}'_k(t_0)/\theta'_l(t_0)) + O(\epsilon). \end{aligned}$$

Since  $\theta'_k(t_0)/\theta'_l(t_0) \leq 1/d < 1 - \Delta$  or  $\theta'_k(t_0)/\theta'_l(t_0) \geq d > 1 + \Delta$  for any  $k \neq l$ , in the first summation of the above equation, we have only one term left,

$$(2.6) \quad a_l(t_0)e^{-i\theta_l(t_0)}\widehat{\psi}(1) = \sum_{k=1}^{\tilde{M}} \tilde{a}_k(t_0)e^{-i\tilde{\theta}_k(t_0)}\widehat{\psi}(\tilde{\theta}'_k(t_0)/\theta'_l(t_0)) + O(\epsilon).$$

Then, there exists at least an  $I(l, t_0) \in \{1, \dots, \tilde{M}\}$  such that

$$(2.7) \quad \left| \widehat{\psi}(\tilde{\theta}'_{I(l, t_0)}(t_0)/\theta'_l(t_0)) \right| > 0,$$

which means that  $1 - \Delta < \tilde{\theta}'_{I(l, t_0)}(t_0)/\theta'_l(t_0) < 1 + \Delta$ . Using the assumption that  $f$  is well separated with frequency ratio  $d$  and  $0 < \Delta < \frac{\sqrt{d}-1}{\sqrt{d}+1}$ , for any  $k \neq l$ , we obtain

$$(2.8) \quad \frac{\theta'_k(t_0)}{\theta'_l(t_0)} \geq d \quad \text{or} \quad \frac{\theta'_k(t_0)}{\theta'_l(t_0)} \leq \frac{1}{d}.$$

This gives that

$$(2.9) \quad \frac{\tilde{\theta}'_{I(l, t_0)}(t_0)}{\theta'_k(t_0)} = \frac{\tilde{\theta}'_{I(l, t_0)}(t_0)}{\theta'_l(t_0)} \cdot \frac{\theta'_l(t_0)}{\theta'_k(t_0)} \geq d(1 - \Delta) > 1 + \Delta$$

or

$$(2.10) \quad \frac{\tilde{\theta}'_{I(l, t_0)}(t_0)}{\theta'_k(t_0)} = \frac{\tilde{\theta}'_{I(l, t_0)}(t_0)}{\theta'_l(t_0)} \cdot \frac{\theta'_l(t_0)}{\theta'_k(t_0)} \leq \frac{1 + \Delta}{d} < 1 - \Delta.$$

Then, for any  $k \neq l$ , we have

$$(2.11) \quad \left| \widehat{\psi}(\tilde{\theta}'_{I(l, t_0)}(t_0)/\theta'_k(t_0)) \right| = 0.$$

This implies that  $I(k, t_0) \neq I(l, t_0)$ ,  $k \neq l$ . Then we get that

$$(2.12) \quad \tilde{M} \geq M.$$

Since  $(\tilde{a}_k, \tilde{\theta}_k)_{1 \leq k \leq \tilde{M}}$  is a solution of  $(P_0)$ , we also have  $\tilde{M} \leq M$ . This implies that

$$(2.13) \quad \tilde{M} = M,$$

and for any  $t \in [0, 1]$ ,  $I(\cdot, t) : \{1, \dots, M\} \rightarrow \{1, \dots, M\}$  is a one-to-one map. Then, we can define its inverse map  $I^{-1}(\cdot, t) : \{1, \dots, M\} \rightarrow \{1, \dots, M\}$ .

For any  $k = 1, \dots, M$ , we study the function  $I^{-1}(k, \cdot) : [0, 1] \rightarrow \{1, \dots, M\}$ . Using the condition that  $\frac{\tilde{\theta}''}{(\tilde{\theta}')^2} \leq \epsilon$  and the signal  $f$  is well separated, it is easy to see that  $I^{-1}(k, \cdot)$  is a constant over  $[0, 1]$ , i.e.,

$$(2.14) \quad I^{-1}(k, t) = I^{-1}(k, 0) \quad \forall t \in [0, 1], \quad k = 1, \dots, M.$$

Otherwise, suppose there exists  $t_0 \in [0, 1]$  such that  $I^{-1}(k, t_0) \neq I^{-1}(k, 0)$ . Let  $A = \{0 \leq t \leq t_0 : I^{-1}(k, t) = I^{-1}(k, 0)\}$  and  $\xi = \sup A$ . Then for any  $\eta > 0$ , there exist  $t_1, t_2 \in [0, 1]$  such that

$$t_1 < \xi < t_2, \quad |t_2 - t_1| < \eta, \quad I^{-1}(k, 0) = I^{-1}(k, t_1) \neq I^{-1}(k, t_2).$$

Denote

$$(2.15) \quad I^{-1}(k, t_1) = k_1, \quad I^{-1}(k, t_2) = k_2, \quad k_1 \neq k_2.$$

Then, by the definition of the map  $I(k, t)$ , we have

$$1 - \Delta < \frac{\tilde{\theta}'_k(t_1)}{\theta'_{k_1}(t_1)} < 1 + \Delta, \quad 1 - \Delta < \frac{\tilde{\theta}'_k(t_2)}{\theta'_{k_2}(t_2)} < 1 + \Delta.$$

Without loss of generality, we assume that  $\theta'_{k_2} > \theta'_{k_1}$ . Then, we have

$$\tilde{\theta}'_k(t_1) < (1 + \Delta)\theta'_{k_1}(t_1), \quad \tilde{\theta}'_k(t_2) > (1 - \Delta)\theta'_{k_2}(t_1).$$

Now, we let  $\eta \rightarrow 0$  and have  $t_1, t_2 \rightarrow \xi$ , which gives

$$\tilde{\theta}'_k(\xi) \leq (1 + \Delta)\theta'_{k_1}(\xi), \quad \tilde{\theta}'_k(\xi) \geq (1 - \Delta)\theta'_{k_2}(\xi).$$

On the other hand, since  $f$  is well separated, we know that

$$\frac{\theta'_{k_1}(t)}{\theta'_{k_2}(t)} \leq 1/d.$$

Then, we have

$$\tilde{\theta}'_k(\xi) \leq (1 + \Delta)\theta'_{k_1}(\xi), \quad \tilde{\theta}'_k(\xi) \geq d(1 - \Delta)\theta'_{k_1}(\xi) > (1 + \Delta)\theta'_{k_1}(\xi),$$

which is a contradiction. This means that  $I^{-1}(k, \cdot)$  is a constant over  $[0, 1]$ , and we can assume

$$(2.16) \quad I^{-1}(k, t) = k \quad \forall t \in [0, 1], \quad k = 1, \dots, M.$$

Now we have that

$$(2.17) \quad 1 - \Delta < \frac{\tilde{\theta}'_k(t)}{\theta'_k(t)} < 1 + \Delta \quad \forall t \in [0, 1], \quad k = 1, \dots, M.$$

Using the assumption that the signal  $f$  is well separated with ratio  $d$  and the choice of  $\psi$ , we have that for any  $k, l = 1, \dots, M$ ,  $k \neq l$ ,

$$(2.18) \quad \hat{\psi}(\omega\theta'_l(t)) = \hat{\psi}(\omega\tilde{\theta}'_l(t)) = 0 \quad \forall (t, \omega) \in U_{\theta_k},$$

$$(2.19) \quad \hat{\psi}(\omega\theta'_l(t)) = \hat{\psi}(\omega\tilde{\theta}'_l(t)) = 0 \quad \forall (t, \omega) \in U_{\tilde{\theta}_k}.$$

On the other hand, using Lemma 2.1 we have

$$(2.20) \quad \begin{aligned} \omega^{-1/2}\mathcal{W}(f)(t, \omega) &= \sum_{k=1}^M a_k(t)e^{-i\theta_k(t)}\hat{\psi}(\omega\theta'_k(t)) + O(\epsilon) \\ &= \sum_{k=1}^M \tilde{a}_k(t)e^{-i\tilde{\theta}_k(t)}\hat{\psi}(\omega\tilde{\theta}'_k(t)) + O(\epsilon). \end{aligned}$$

Using the three relations (2.18), (2.19), and (2.20), we have

$$(2.21) \quad \left| a_k(t)e^{-i\theta_k(t)}\hat{\psi}(\omega\theta'_k(t)) - \tilde{a}_k(t)e^{-i\tilde{\theta}_k(t)}\hat{\psi}(\omega\tilde{\theta}'_k(t)) \right| = O(\epsilon) \quad \forall (t, \omega) \in U_{\theta_k} \cup U_{\tilde{\theta}_k},$$

which implies that

$$(2.22) \quad \left| a_k(t) \widehat{\psi}(\omega \theta'_k(t)) \right| - \left| \tilde{a}_k(t) \widehat{\psi}(\omega \tilde{\theta}'_k(t)) \right| = O(\epsilon) \quad \forall (t, \omega) \in U_{\theta_k} \bigcup U_{\tilde{\theta}_k}.$$

Next, we will prove that  $|a_k(t) - \tilde{a}_k(t)| = O(\epsilon)$  and  $|\theta'_k(t) - \tilde{\theta}'_k(t)| = O(\epsilon)$  from (2.22). First, we consider the envelopes  $a_k$  and  $\tilde{a}_k$ .

If  $a_k(t) > \tilde{a}_k(t)$ , we choose  $\omega = 1/\theta'_k(t)$  to get

$$(2.23) \quad a_k(t) \left| \widehat{\psi}(1) \right| - \tilde{a}_k(t) \left| \widehat{\psi}(\tilde{\theta}'_k(t)/\theta'_k(t)) \right| = O(\epsilon),$$

where we have used the fact that  $|\widehat{\psi}(\xi)|$  achieves its maximum at  $\xi = 1$ . Since  $a_k(t) > \tilde{a}_k(t)$ , we have

$$(2.24) \quad 0 \leq \left| \widehat{\psi}(1) \right| (a_k(t) - \tilde{a}_k(t)) \leq a_k(t) \left| \widehat{\psi}(1) \right| - \tilde{a}_k(t) \left| \widehat{\psi}(\tilde{\theta}'_k(t)/\theta'_k(t)) \right| = O(\epsilon).$$

This proves that

$$(2.25) \quad a_k(t) - \tilde{a}_k(t) = O(\epsilon).$$

If  $a_k(t) < \tilde{a}_k(t)$ , we take  $\omega = 1/\tilde{\theta}'_k(t)$ . By following a similar argument, we can prove

$$(2.26) \quad \tilde{a}_k(t) - a_k(t) = O(\epsilon).$$

Combining these two cases, we obtain

$$(2.27) \quad |a_k(t) - \tilde{a}_k(t)| = O(\epsilon).$$

Substituting the above relation in (2.22), we get

$$(2.28) \quad \left| \widehat{\psi}(\omega \theta'_k(t)) \right| - \left| \widehat{\psi}(\omega \tilde{\theta}'_k(t)) \right| = O(\epsilon) \quad \forall (t, \omega) \in U_{\theta_k} \bigcup U_{\tilde{\theta}_k}.$$

For any  $t \in [0, 1]$ , let  $\omega = (1 - \Delta/2)/\theta'_k(t)$ ; then we have

$$(2.29) \quad \left| \widehat{\psi}(1 - \Delta/2) \right| - \left| \widehat{\psi}[(1 - \Delta/2)\tilde{\theta}'_k(t)/\theta'_k(t)] \right| = O(\epsilon).$$

Since  $\widehat{\psi}$  is a fifth order B-Spline function and  $\epsilon \ll 1$ , it is easy to see that there exists a constant  $C > 0$  such that

$$(2.30) \quad \left| (1 - \Delta/2) \left( \tilde{\theta}'_k(t)/\theta'_k(t) - 1 \right) \right| \leq C\epsilon.$$

Since  $\widehat{\psi}(\omega)$  is a fifth order B-Spline function,  $\widehat{\psi}(\omega)$  is symmetric with respect to  $\omega = 1$ . Thus we have  $\widehat{\psi}(1 - \Delta/2) = \widehat{\psi}(1 + \Delta/2)$ . Then, there is also another possibility:

$$(2.31) \quad \left| (1 - \Delta/2)\tilde{\theta}'_k(t)/\theta'_k(t) - (1 + \Delta/2) \right| \leq C\epsilon.$$

If (2.30) holds, then we get

$$(2.32) \quad \frac{|\theta'_k(t) - \tilde{\theta}'_k(t)|}{\theta'_k(t)} = O(\epsilon).$$



If (2.31) holds, then we have

$$(2.33) \quad \frac{\tilde{\theta}'_k(t)}{\theta'_k(t)} \geq \frac{1 + \Delta/2 - C\epsilon}{1 - \Delta/2} \geq 1 + \Delta,$$

where we have used the assumption that  $\epsilon \ll 1$ .

Then, letting  $\omega = 1/\theta'_k(t)$  in (2.28), we have

$$(2.34) \quad 1 = \left| \widehat{\psi}(1) \right| - \left| \widehat{\psi} \left( \tilde{\theta}'_k(t)/\theta'_k(t) \right) \right| = O(\epsilon).$$

This argument shows that (2.31) cannot be true. This completes the proof.  $\square$

**2.2. Discussion on signals with close frequencies.** In this section, we will give a brief discussion on the signal with close frequencies, i.e., the frequency ratio  $d \rightarrow 1$  in the definition of a well-separated signal.

In the proof of Theorem 1.3, the frequency ratio  $d$  seems to be arbitrary, as long as it is larger than 1. Actually, the distance between  $d$  and 1 cannot be too small. The gap is determined by the separation factor  $\epsilon$ . First, it is easy to show that the integrals  $I_1, I_2, I_3$  in Lemma 2.1 satisfy

$$(2.35) \quad I_1 = O(\Delta^{-1}), \quad I_2 = O(\Delta^{-2}), \quad I_3 = O(\Delta^{-3}).$$

When  $d$  approaches 1,  $\Delta = \frac{\sqrt{d}-1}{\sqrt{d}+1}$  becomes smaller; then  $I_1, I_2, I_3$  become larger. When  $I_1, I_2, I_3$  is as large as  $1/\epsilon$ , Lemma 2.1 breaks down, which in turn leads to the breakdown of the proof of Theorem 1.3.

On the other hand, the frequency ratio is defined pointwise. As long as there exists a point such that  $d-1 \gg \epsilon$ , where  $d$  is the frequency ratio, the argument before (2.13) in the proof of Theorem 1.3 still applies, which means that the decomposition given in (1.6) is still an optimal solution of the optimization problem  $(P_0)$ . But in this case, the uniqueness is not guaranteed. To illustrate this point, we construct an example such that the solution of  $(P_0)$  is not unique:

$$(2.36) \quad f(t) = \cos \theta_1(t) + \cos \theta_2(t), \quad t \in [0, 1],$$

$$(2.37) \quad \theta_1(t) = 6\pi kt + k\pi, \quad \theta_2(t) = 8\pi kt + k \sin 2\pi t,$$

where  $k$  is a positive integer. We can choose  $k$  large enough such that  $\cos \theta_1, \cos \theta_2 \in U_\epsilon$ . At  $t = 0$ ,  $\theta'_2(0)/\theta'_1(0) = 4/3$ . From the discussion above, we know that (2.36) gives a solution of  $(P_0)$ . On the other hand, it is easy to check that the following decomposition is also a solution of  $(P_0)$ :

$$(2.38) \quad f(t) = \cos \phi_1(t) + \cos \phi_2(t), \quad t \in [0, 1],$$

$$\phi_1(t) = \begin{cases} 6\pi kt + k\pi, & t \in [0, 1/2], \\ 8\pi kt + k \sin 2\pi t, & t \in (1/2, 1], \end{cases} \quad \phi_2(t) = \begin{cases} 8\pi kt + k \sin 2\pi t, & t \in [0, 1/2], \\ 6\pi kt + k\pi, & t \in (1/2, 1]. \end{cases}$$

This example shows that when the frequencies intersect, the solution of  $(P_0)$  may not be unique. In this case, we need to impose some extra constraints to obtain uniqueness of the decomposition. One natural idea is to pick the solution according to the regularity of  $a(t)$  and  $\theta'(t)$ . The solution with smoother amplitude and frequency is favorable. One method based on this idea is proposed in [12] to decompose signals that do not have well-separated IMFs. In that method, the regularity is related with the sparsity over the Fourier (or wavelet) dictionary. In the above example, this method would prefer the decomposition (2.36), since in this decomposition, both the amplitude and the frequencies are very sparse over the Fourier dictionary.

**3. Optimal solutions to the (P<sub>2</sub>) problem.** Using the result in the previous section, we know that for signals that are well separated, the solution of (P<sub>0</sub>) is unique up to the separation factor  $\epsilon$ . One natural question is how to find this unique solution.

Based on matching pursuit, we propose the following algorithm to solve (P<sub>0</sub>) approximately.

- $r_0 = f, \quad k = 1.$

Step 1: Solve the following nonlinear least-squares problem:

$$(3.1) \quad \begin{aligned} (a_k, \theta_k) \in & \underset{a, \theta}{\operatorname{argmin}} \quad \|r_{k-1} - a \cos \theta\|_{l^2}^2 \\ & \text{subject to} \quad a, \theta \in U_\epsilon. \end{aligned}$$

Step 2: Update the residual

$$(3.2) \quad r_k = f - \sum_{j=1}^k a_j \cos \theta_j.$$

Step 3: If  $\|r_k\|_{l^2} < \epsilon_0$ , stop. Otherwise, set  $k = k + 1$  and go to Step 1.

In the above algorithm based on matching pursuit, each IMF is given by solving the following nonlinear least-squares problem:

$$(P_2) \quad \begin{aligned} & \text{Minimize} \quad p(a, \theta) := \|f(t) - a(t) \cos \theta(t)\|_{L^2}^2 \\ & \text{subject to} \quad (a, \theta) \in U_\epsilon. \end{aligned}$$

We would like to know under what conditions on each IMF  $a_k(t) \cos \theta_k(t)$  of  $f(t)$  the above nonlinear least-squares problem could provide a local (approximate) optimizer to the (P<sub>0</sub>) problem.

In the computations, we always deal with signals with finite time span. Without loss of generality, in this section we assume the time span of the signal is  $[0, 1]$ . For the signal with finite length, introducing larger errors near the end points of the time interval is unavoidable. This is also known as the “end effect.” To simplify the analysis, we assume that the signal is periodic over  $[0, 1]$ . For those signals that are not periodic, we first multiply a cutoff function to make the signal vanish near the boundary and then treat it as a periodic signal. This means that the analysis in this section is valid only in the interior region away from the boundary.

For a periodic signal, we prove that each IMF is a local minimizer of (P<sub>2</sub>) as long as the signal satisfies some assumptions. This result is stated in Theorem 3.1.

**THEOREM 3.1.** *Let  $f(t)$  be a function satisfying the scale separation property with separation factor  $\epsilon$  and frequency ratio  $d$  as defined in Definition 1.2:*

$$f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t) + r(t), \quad a_k \cos \theta_k \in U_\epsilon, \quad a_k = O(1), \quad r = O(\epsilon).$$

*Suppose there exist  $\alpha \in [1, d]$  and  $l \in \{1, \dots, M\}$  such that*

$$(3.3) \quad \alpha^{-1} \theta'_l(t) \leq \theta'(t) \leq \alpha \theta'_l(t) \quad \forall t \in [0, 1].$$

*If*

$$(3.4) \quad p(a, \theta) \leq p(a_l, \theta_l),$$

where  $p(a, \theta)$  is given in  $(P_2)$ , then we have

$$(3.5) \quad \frac{\|a \cos \theta - a_l \cos \theta_l\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} = O(\sqrt{\epsilon}).$$

*Proof.* First we know

$$(3.6) \quad \begin{aligned} 0 &\geq p(a, \theta) - p(a_l, \theta_l) \\ &= \|f(t) - a(t) \cos \theta(t)\|_{L^2}^2 - \|f(t) - a_l(t) \cos \theta_l(t)\|_{L^2}^2 \\ &= \left\| \sum_{k \neq l} a_k \cos \theta_k + r(t) + a_l(t) \cos \theta_l(t) - a(t) \cos \theta(t) \right\|_{L^2}^2 - \left\| \sum_{k \neq l} a_k \cos \theta_k + r(t) \right\|_{L^2}^2 \\ &= \|a_l \cos \theta_l - a \cos \theta\|_{L^2}^2 + 2 \left\langle a_l \cos \theta_l - a \cos \theta, \sum_{k \neq l} a_k \cos \theta_k + r(t) \right\rangle, \end{aligned}$$

where the first equality follows from the definition of  $p(a, \theta)$  in  $(P_2)$ . In the rest of the proof, we try to control the second term of the above inequality.

It is easy to verify that

$$(3.7) \quad \left| \left\langle a_l \cos \theta_l, \sum_{k \neq l} a_k \cos \theta_k \right\rangle \right| \leq \left[ \sum_{k \neq l} \mu_{k,l} \|a_l \cos \theta_l\|_{L^2} \|a_k \cos \theta_k\|_{L^2} \right] = \delta_1 \|a_l \cos \theta_l\|_{L^2}^2,$$

where

$$(3.8) \quad \mu_{k,l} = \frac{|\langle a_l \cos \theta_l, a_k \cos \theta_k \rangle|}{\|a_l \cos \theta_l\|_{L^2} \|a_k \cos \theta_k\|_{L^2}}, \quad \delta_1 = \sum_{k \neq l} \mu_{k,l} \frac{\|a_k \cos \theta_k\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}}.$$

Similarly, we have

$$(3.9) \quad \begin{aligned} &\left| \left\langle a \cos \theta, \sum_{k \neq l} a_k \cos \theta_k + r(t) \right\rangle \right| \\ &< \delta_2 \|a \cos \theta\|_{L^2} \|a_l \cos \theta_l\|_{L^2} \\ &\leq \delta_2 \|a \cos \theta - a_l \cos \theta_l\|_{L^2} \cdot \|a_l \cos \theta_l\|_{L^2} + \delta_2 \|a_l \cos \theta_l\|_{L^2}^2, \end{aligned}$$

with

$$(3.10) \quad \begin{aligned} \delta_2 &= \sum_{k \neq l} \mu_{k,l,\alpha} \frac{\|a_k \cos \theta_k\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} + \frac{\|r(t)\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}}, \\ \mu_{k,l,\alpha} &= \frac{|\langle a \cos \theta, a_k \cos \theta_k \rangle|}{\|a \cos \theta\|_{L^2} \|a_k \cos \theta_k\|_{L^2}}. \end{aligned}$$

Thus it follows from (3.6), (3.7), and (3.9) that

$$(3.11) \quad \begin{aligned} 0 &> \|a_l \cos \theta_l - a \cos \theta\|_{L^2}^2 - 2\delta_2 \|a \cos \theta - a_l \cos \theta_l\|_{L^2} \\ &\quad \cdot \|a_l \cos \theta_l\|_{L^2} - 2(\delta_1 + \delta_2) \|a_l \cos \theta_l\|_{L^2}^2, \end{aligned}$$

which implies

$$(3.12) \quad \frac{\|a \cos \theta - a_l \cos \theta_l\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} \leq \delta_2 + \sqrt{\delta_2^2 + 2(\delta_1 + \delta_2)}.$$

Here,  $\mu_{k,l}$  and  $\mu_{k,l,\alpha}$  are just the coherences between  $a_l \cos \theta_l$ ,  $a \cos \theta$ , and  $a_k \cos \theta_k$ . Under the assumption of scale separation and the assumption that different IMFs are well separated, the behavior of  $a_l \cos \theta_l$ ,  $a \cos \theta$ , and  $a_k \cos \theta_k$  is close to that of the standard Fourier basis. Then, it is natural to expect that the coherences,  $\mu_{k,l}$  and  $\mu_{k,l,\alpha}$ , are small. Actually, the smallness of  $\mu_{k,l}$  and  $\mu_{k,l,\alpha}$  is given in Corollary 3.3. In particular, the estimate (3.20) from Corollary 3.3 shows that for all  $k \neq l$  we have

$$(3.13) \quad \mu_{k,l} = O(\epsilon), \quad \mu_{k,l,\alpha} = O(\epsilon).$$

Together with the assumption that  $r(t) = O(\epsilon)$ , we get

$$(3.14) \quad \delta_1 = O(\epsilon), \quad \delta_2 = O(\epsilon).$$

It follows from (3.12) that

$$(3.15) \quad \frac{\|a \cos \theta - a_l \cos \theta_l\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} = O(\sqrt{\epsilon}).$$

This completes the proof.  $\square$

In the proof of the above theorem, we have used the estimate (3.13) for  $\mu_{k,l}$  and  $\mu_{k,l,\alpha}$ . This estimate can be derived by the following lemma.

LEMMA 3.2. *Let  $(a, \theta) \in U_\epsilon$  be such that  $a \cos \theta$  has period 1. Then we have*

$$(3.16) \quad \left(\frac{1}{2} - 3\epsilon\right) \|a(t)\|_{L^2}^2 \leq \|a(t) \cos \theta(t)\|_{L^2}^2 \leq \left(\frac{1}{2} + 3\epsilon\right) \|a(t)\|_{L^2}^2.$$

Furthermore, if there is another pair  $(\bar{a}, \bar{\theta}) \in U_\epsilon$  being periodic over  $[0, 1]$  such that

$$(3.17) \quad \beta := \min_{t \in [0,1]} \frac{\bar{\theta}'(t)}{\theta'(t)} > 1,$$

then we have

$$(3.18) \quad |\langle a \cos \theta, \bar{a} \cos \bar{\theta} \rangle| < 4\epsilon \left(1 + \frac{1}{(1 - \beta^{-1})^2}\right) \int_0^1 a(t) \bar{a}(t) dt.$$

The proof of this lemma can be found in Appendix B. Then (3.13) is just a direct corollary of the above lemma.

COROLLARY 3.3. *Let  $(a_k, \theta_k)$ ,  $k = 1, \dots, M$ , be well separated with frequency ratio  $d$  and separation factor  $\epsilon$  as defined in Definition 1.2. Let  $(a, \theta) \in U_\epsilon$ , and let there exist  $\alpha \in [1, d)$  and  $l \in \{1, \dots, M\}$  such that*

$$(3.19) \quad \alpha^{-1} \theta'_l(t) \leq \theta'(t) \leq \alpha \theta'_l(t) \quad \forall t \in [0, 1].$$

Then for any  $1 \leq k \neq l \leq M$ , we have

$$(3.20) \quad \begin{aligned} \mu_{k,l} &:= \frac{|\langle a_k \cos \theta_k, a_l \cos \theta_l \rangle|}{\|a_l \cos \theta_l\|_{L^2} \|a_k \cos \theta_k\|_{L^2}} = O(\epsilon), \\ \mu_{k,l,\alpha} &:= \frac{|\langle a_k \cos \theta_k, a \cos \theta \rangle|}{\|a \cos \theta\|_{L^2} \|a_k \cos \theta_k\|_{L^2}} = O(\epsilon). \end{aligned}$$

*Proof.* For any  $1 \leq k \neq l \leq M$ , suppose  $\theta'_k > \theta'_l$ . Since  $(a_k, \theta_k)$ ,  $k = 1, \dots, M$ , are well separated, we have

$$(3.21) \quad \min_{t \in [0,1]} \frac{\theta'_k}{\theta'_l} > d^{|k-l|}.$$

Using Lemma 3.2, it is easy to check that

$$\begin{aligned} \mu_{k,l} &= \frac{|\langle a_k \cos \theta_k, a_l \cos \theta_l \rangle|}{\langle a_k, a_l \rangle} \cdot \frac{\langle a_k, a_l \rangle}{\|a_l \cos \theta_l\|_{L^2} \|a_k \cos \theta_k\|_{L^2}} \\ &\leq \frac{|\langle a_k \cos \theta_k, a_l \cos \theta_l \rangle|}{\langle a_k, a_l \rangle} \cdot \frac{\|a_k\|_{L^2}}{\|a_k \cos \theta_k\|_{L^2}} \cdot \frac{\|a_l\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} \\ &< 4\epsilon \left( \frac{1}{2} - 3\epsilon \right)^{-1} \left( 1 + \frac{1}{(1 - d^{-|l-k|})^2} \right). \end{aligned}$$

For any  $1 \leq k \neq l \leq M$ , since  $(a_k, \theta_k)$ ,  $k = 1, \dots, M$ , are well separated and

$$(3.22) \quad \alpha^{-1} \theta'_l(t) \leq \theta'(t) \leq \alpha \theta'_l(t) \quad \forall t \in [0, 1],$$

if  $\theta'_k < \theta'_l$ , we know that

$$(3.23) \quad \frac{\theta'}{\theta'_k} \geq \alpha^{-1} d^{l-k}.$$

In addition, if  $\theta'_k > \theta'_l$ , we have

$$(3.24) \quad \frac{\theta'_k}{\theta'} \geq \alpha^{-1} d^{k-l}.$$

Then it follows from Lemma 3.2 that, for  $k \neq l$ ,

$$\begin{aligned} \mu_{k,l,\alpha} &\leq \frac{|\langle a_k \cos \theta_k, a \cos \theta \rangle|}{\langle a_k, a \rangle} \cdot \frac{\|a_k\|_{L^2}}{\|a_k \cos \theta_k\|_{L^2}} \cdot \frac{\|a\|_{L^2}}{\|a \cos \theta\|_{L^2}} \\ &< 4\epsilon \left( \frac{1}{2} - 3\epsilon \right)^{-1} \left( 1 + \frac{1}{(1 - \alpha d^{-|l-k|})^2} \right). \quad \square \end{aligned}$$

*Remark 3.1.* We remark that in Corollary 3.3 and Theorem 3.1 the condition (3.3) could be replaced by

$$(3.25) \quad \frac{\theta'(t)}{\theta'_{l-1}(t)} \geq \frac{d}{\alpha}, \quad \frac{\theta'_{l+1}(t)}{\theta'(t)} \geq \frac{d}{\alpha} \quad \forall t \in [0, 1].$$

If we put a stronger assumption on the separation ratio  $d$  such that  $d > M'^2$  (where  $M'$  is defined in Definition 1.1), then we obtain the following theorem about the global minimizer of the  $(P_2)$  problem.

**THEOREM 3.4.** *Let  $f(t)$  be a function satisfying the scale separation property with separation factor  $\epsilon$  and frequency ratio  $d$  as defined in Definition 1.2:*

$$f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t) + r(t), \quad a_k \cos \theta_k \in U_\epsilon, \quad a_k = O(1), \quad r = O(\epsilon).$$

Suppose further that  $d > M^2$ , and that there exists  $l \in \overline{1, M}$  such that

$$(3.26) \quad \|a_k \cos \theta_k\| < \|a_l \cos \theta_l\| \quad \forall k \in \overline{1, M}, \quad k \neq l.$$

If

$$(3.27) \quad p(a, \theta) \leq p(a_l, \theta_l),$$

where  $p(a, \theta)$  is given in  $(P_2)$ , then we have

$$(3.28) \quad \frac{\|a \cos \theta - a_l \cos \theta_l\|_{L^2}}{\|a_l \cos \theta_l\|_{L^2}} = O(\sqrt{\epsilon}).$$

*Proof.* First, we claim that, for each  $k \in \overline{1, M}$ ,

$$(3.29) \quad p(a_k, \theta_k) = \|f\|_{L^2}^2 - \|a_k \cos \theta_k\|_{L^2}^2 + O(\epsilon).$$

Granting this, it follows from (3.26) that

$$(3.30) \quad p(a_k, \theta_k) > p(a_l, \theta_l) \quad \forall k \neq l.$$

To show (3.29), we notice that

$$(3.31) \quad \begin{aligned} p(a_k, \theta_k) &= \|f - a_k \cos \theta_k\|_{L^2}^2 \\ &= \|f\|_{L^2}^2 + \|a_k \cos \theta_k\|_{L^2}^2 - 2 \left\langle a_k \cos \theta_k + \sum_{k' \neq k} a_{k'} \cos \theta_{k'} + r(t), a_k \cos \theta_k \right\rangle \\ &= \|f\|_{L^2}^2 - \|a_k \cos \theta_k\|_{L^2}^2 - 2 \left\langle \sum_{k' \neq k} a_{k'} \cos \theta_{k'} + r(t), a_k \cos \theta_k \right\rangle. \end{aligned}$$

By Lemma 3.2, we get

$$(3.32) \quad \langle a_{k'} \cos \theta_{k'}, a_k \cos \theta_k \rangle = O(\epsilon)$$

for each  $k' \neq k$ . By the assumption in Definition 1.2, we have  $r(t) = O(\epsilon)$ , which implies

$$(3.33) \quad \langle r(t), a_k \cos \theta_k \rangle = O(\epsilon).$$

Substituting (3.32) and (3.33) into (3.31) proves (3.29).

Let  $\alpha := \frac{\sqrt{d}}{M'} (so \alpha > 1)$ , and let

$$(3.34) \quad \alpha_k := \begin{cases} \max_{t \in [0,1]} \frac{\theta'(t)}{\theta'_k(t)}, & k = 1, \\ \max_{t \in [0,1]} \frac{\theta'_k(t)}{\theta'(t)}, & k = M, \\ \max \left\{ \max_{t \in [0,1]} \frac{\theta'(t)}{\theta'_k(t)}, \max_{t \in [0,1]} \frac{\theta'_k(t)}{\theta'(t)} \right\}, & 1 < k < M, \end{cases}$$

$$(3.35) \quad \alpha_{min} := \min_{1 \leq k \leq M} \{\alpha_k\}.$$

Let  $k_0$  be such that  $\alpha_{k_0} = \alpha_{min}$ . We claim the following:

(1) If  $k_0 > 1$ ,

$$(3.36) \quad \frac{\theta'(t)}{\theta'_{k_0-1}(t)} \geq \alpha \quad \forall t \in [0, 1].$$

(2) If  $k_0 < K$ ,

$$(3.37) \quad \frac{\theta'_{k_0+1}(t)}{\theta'(t)} \geq \alpha \quad \forall t \in [0, 1].$$

We prove item (1) by contradiction. Assume that there exists  $t_1 \in [0, 1]$  such that

$$(3.38) \quad \frac{\theta'(t_1)}{\theta'_{k_0-1}(t_1)} < \alpha.$$

Then for any  $t \in [0, 1]$ , we have

$$(3.39) \quad \frac{\theta'(t)}{\theta'_{k_0-1}(t)} = \frac{\theta'(t)}{\theta'(t_1)} \cdot \frac{\theta'(t_1)}{\theta'_{k_0-1}(t_1)} \cdot \frac{\theta'_{k_0-1}(t_1)}{\theta'_{k_0-1}(t)} < M'^2 \alpha = M' \sqrt{d}.$$

On the other hand, (3.38) also implies that

$$(3.40) \quad \frac{\theta'_{k_0}(t_1)}{\theta'(t_1)} = \frac{\theta'_{k_0}(t_1)}{\theta'_{k_0-1}(t_1)} \left( \frac{\theta'(t_1)}{\theta'_{k_0-1}(t_1)} \right)^{-1} > \frac{d}{\alpha} = M' \sqrt{d}.$$

Using (3.39), (3.40), and the estimates

$$(3.41) \quad \frac{\theta'_{k_0-1}(t_1)}{\theta'(t_1)} \leq \frac{1}{d} \cdot \frac{\theta'_{k_0}(t_1)}{\theta'(t_1)} < \frac{\theta'_{k_0}(t_1)}{\theta'(t_1)},$$

we get

$$(3.42) \quad \alpha_{k_0-1} < \max_{t \in [0, 1]} \frac{\theta'_{k_0}(t)}{\theta'(t)} \leq \alpha_{k_0}.$$

This contradicts  $\alpha_{k_0} = \alpha_{\min}$ . So item (1) is satisfied. Similarly, we can show that item (2) is true.

It follows from (3.27) and (3.30) that

$$(3.43) \quad p(a, \theta) \leq p(a_{k_0}, \theta_{k_0}).$$

Using Theorem 3.1 by replacing condition (3.3) by (3.25) in Remark 3.1, we obtain

$$(3.44) \quad \frac{\|a \cos \theta - a_{k_0} \cos \theta_{k_0}\|_{L^2}}{\|a_{k_0} \cos \theta_{k_0}\|_{L^2}} = O(\sqrt{\epsilon}).$$

This implies

$$(3.45) \quad \|a \cos \theta - a_{k_0} \cos \theta_{k_0}\|_{L^2} = O(\sqrt{\epsilon}).$$

Therefore we obtain

$$(3.46) \quad \begin{aligned} & |p(a, \theta) - p(a_{k_0}, \theta_{k_0})| \\ &= \|a_{k_0} \cos \theta_{k_0} - a \cos \theta\|_{L^2}^2 + 2 \left| \left\langle a_{k_0} \cos \theta_{k_0} - a \cos \theta, \sum_{k \neq k_0} a_k \cos \theta_k + r(t) \right\rangle \right| \\ &= O(\sqrt{\epsilon}), \end{aligned}$$

where the first equality is deduced using an argument similar to the equalities in (3.6), and the second one follows from (3.45). So by (3.27), we see that  $k_0 = l$  and  $\alpha_k > \alpha_l$  for all  $k \neq l$ . Thus (3.44) implies (3.28). This completes the proof.  $\square$

The condition that  $d > M'^2$  is very strong when  $M'$  is big. However, if this condition is violated, the global minimizer to the  $(P_2)$  problem may not be any of the IMFs in the sparsest representation. For example, let us consider an artificial signal  $f(t)$  defined as follows:

(3.47)

$$f(t) = a_1(t) \cos \theta_1(t) + a_2(t) \cos \theta_2(t), \quad t \in [0, 6],$$

$$\text{where } a_1(t) = 2 + t, \quad \theta_1(t) = \begin{cases} 10\pi t, & t \in [0, 2], \\ 20\pi + 10\pi(t-2) + \frac{5\pi}{3}(t-2)^3, & t \in [2, 3], \\ 50\pi + 20\pi(t-4) - \frac{5\pi}{3}(t-4)^3, & t \in [3, 4], \\ 50\pi + 20\pi(t-4), & t \in [4, 6], \end{cases}$$

$$a_2(t) = 8 - t, \quad \theta_2(t) = 2\theta_1(t).$$

Here we see that  $d = M' = 2$ . Moreover, it is easy to verify that

(3.48)

$$\left| \frac{a'_1(t)}{\theta'_1(t)} \right| \leq \frac{1}{10\pi}, \quad \left| \frac{\theta''_1(t)}{(\theta'_1(t))^2} \right| \leq \frac{1}{10\pi}; \quad \left| \frac{a'_2(t)}{\theta'_2(t)} \right| \leq \frac{1}{20\pi}, \quad \left| \frac{\theta''_2(t)}{(\theta'_2(t))^2} \right| \leq \frac{1}{20\pi} \quad \forall t \in [0, 6].$$

So  $a_1(t) \cos \theta_1(t)$  and  $a_2(t) \cos \theta_2(t)$  are well separated and both satisfy the scale separation property on the time domain  $[0, 6]$ . If we attempt to solve problem  $(P_2)$  with parameter  $\epsilon = \frac{1}{10\pi}$ , we will find that the solution  $(a, \theta)$  with

$$(3.49) \quad a(t) = \max\{a_1(t), a_2(t)\} = 5 + |t - 3|, \quad \theta(t) = 20\pi t$$

is better than  $(a_1, \theta_1)$  or  $(a_2, \theta_2)$  (here we note that  $(a, \theta)$  is obtained by simply connecting the left part of  $(a_2, \theta_2)$  and the right part of  $(a_1, \theta_1)$ ). In fact, numerical estimation gives us that

$$(3.50) \quad \begin{aligned} p(a, \theta) &= \|f(t) - a(t) \cos \theta(t)\|_{L^2} \approx 72.4, \\ p(a_1, \theta_1) &= \|a_2 \cos \theta_2\|_{L^2}^2 \approx 84, \\ p(a_2, \theta_2) &= \|a_1 \cos \theta_1\|_{L^2}^2 \approx 84. \end{aligned}$$

Thus, if  $d \leq M'^2$ , mode mixing may occur.

However, Theorem 3.4 is still helpful for us even if the condition that  $d > M'^2$  is violated. Consider the signal  $f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t) + r(t)$  on  $[0, 1]$ . We could partition the whole time domain  $[0, 1]$  into  $m$  subintervals  $[t_{i-1}, t_i]$  ( $i = 1, \dots, m$ ), where  $0 = t_0 < t_1 < \dots < t_m = 1$ , and on each subinterval, we would have

$$(3.51) \quad \frac{\sup_{t \in [t_{i-1}, t_i]} \theta'_k(t)}{\inf_{t \in [t_{i-1}, t_i]} \theta'_k(t)} < \sqrt{d}, \quad k = 1, 2, \dots, M.$$

So we could find each IMF on each subinterval efficiently and connect different parts of each IMF. This is the basic idea of the upcoming work of the authors: a two-level method in sparse time-frequency representation of multiscale data.



**4. Concluding remarks.** In this paper, we discussed the uniqueness of the decomposition obtained by the sparse time-frequency decomposition. We proved that under the assumption of scale separation, the decomposition is unique up to an error associated with the scale separation. Moreover, we showed that under the same assumption, nonlinear matching pursuit could be used to obtain this sparse decomposition. The results in this paper establish a solid foundation for the sparse time-frequency decomposition.

In our future work, we would like to relax some of the assumptions of scale separation. In many problems, the decompositions seem to be unique although the scale separation is not satisfied. We plan to conduct further theoretical study and seek guidance on decomposing signals with poor scale separation.

#### Appendix A. Proof of Lemma 2.1.

*Proof of Lemma 2.1.* First, we have that

$$(A.1) \quad \widehat{\psi}(\omega\theta'(t)) = \int_{\mathbb{R}} e^{-i\omega\theta'(t)z} \psi(z) dz = \frac{1}{\omega} \int_{\mathbb{R}} e^{-i\theta'(t)z} \psi\left(\frac{z}{\omega}\right) dz.$$

Using this relation, we obtain

$$(A.2) \quad \begin{aligned} & \frac{1}{\sqrt{\omega}} \int_{\mathbb{R}} a(\tau) e^{-i\theta(\tau)} \psi\left(\frac{\tau-t}{\omega}\right) d\tau - |\omega|^{1/2} a(t) e^{-i\theta(t)} \widehat{\psi}(\omega\theta'(t)) \\ &= \frac{1}{\sqrt{\omega}} \left[ \int_{\mathbb{R}} (a(\tau) - a(t)) e^{-i\theta(\tau)} \psi\left(\frac{\tau-t}{\omega}\right) d\tau \right. \\ & \quad \left. + a(t) \int_{\mathbb{R}} (e^{-i\theta(\tau)} - e^{-i(\theta(t)+\theta'(t)(\tau-t))}) \psi\left(\frac{\tau-t}{\omega}\right) d\tau \right]. \end{aligned}$$

For the first term, we have

$$\begin{aligned} & |\omega|^{-1/2} \int_{\mathbb{R}} (a(\tau) - a(t)) e^{-i\theta(\tau)} \psi\left(\frac{\tau-t}{\omega}\right) d\tau \\ &= |\omega|^{-1/2} \int_{\mathbb{R}} h(\tau, t) e^{-i\theta(\tau)} d\tau \\ &= -i|\omega|^{-1/2} \int_{\mathbb{R}} \left( \frac{h(\tau, t)}{\theta'(\tau)} \right)' e^{-i\theta(\tau)} d\tau, \end{aligned}$$

where

$$(A.3) \quad h(\tau, t) = (a(\tau) - a(t)) \psi\left(\frac{\tau-t}{\omega}\right).$$

Direct calculation gives that

$$(A.4) \quad \left( \frac{h(\tau, t)}{\theta'(\tau)} \right)' = \frac{a'(\tau)}{\theta'(\tau)} \psi\left(\frac{\tau-t}{\omega}\right) + \frac{a(\tau) - a(t)}{\theta'(\tau)} \frac{1}{\omega} \psi'\left(\frac{\tau-t}{\omega}\right) - \frac{h(\tau, t) \theta''(\tau)}{(\theta'(\tau))^2}.$$

Using the assumption that  $(a, \theta) \in U_\epsilon$ , we have

$$(A.5) \quad \frac{a'(\tau)}{\theta'(\tau)} \leq \epsilon, \quad \frac{\theta''(\tau)}{(\theta'(\tau))^2} \leq \epsilon,$$

and

$$(A.6) \quad \left| \frac{a(\tau) - a(t)}{\theta'(\tau)} \right| = \left| \frac{a'(t_\tau)(\tau-t)}{\theta'(\tau)} \right| = \left| \frac{\theta'(t_\tau)}{\theta'(\tau)} \cdot \frac{a'(t_\tau)(\tau-t)}{\theta'(t_\tau)} \right| \leq M' \epsilon |\tau - t|,$$

where  $t_\tau$  is a point between  $t$  and  $\tau$ . Then, we obtain

$$(A.7) \quad |\omega|^{-1/2} \left| \int_{\mathbb{R}} (a(\tau) - a(t)) e^{-i\theta(\tau)} \psi \left( \frac{\tau - t}{\omega} \right) d\tau \right| \leq \epsilon |\omega|^{1/2} [(A + |a(t)| + 1)I_1 + M'I_2],$$

where  $A = \sup_{t \in \mathbb{R}} |a(t)|$  and

$$(A.8) \quad I_1 = \int_{\mathbb{R}} |\psi(\tau)| d\tau, \quad I_2 = \int_{\mathbb{R}} |\tau \psi'(\tau)| d\tau.$$

Now, we turn to bounding the second term of (A.2). We have

$$\begin{aligned} & |\omega|^{-1/2} \int_{\mathbb{R}} (e^{-i\theta(\tau)} - e^{-i(\theta(t) + \theta'(t)(\tau - t))}) \psi \left( \frac{\tau - t}{\omega} \right) d\tau \\ &= |\omega|^{-1/2} \int_{\mathbb{R}} g(\tau, t) e^{-i\theta(\tau)} d\tau \\ &= -i|\omega|^{-1/2} \int_{\mathbb{R}} \left( \frac{g(\tau, t)}{\theta'(\tau)} \right)' e^{-i\theta(\tau)} d\tau, \end{aligned}$$

where

$$(A.9) \quad g(\tau, t) = (1 - e^{i\Delta\theta}) \psi \left( \frac{\tau - t}{\omega} \right)$$

and  $\Delta\theta = \theta(\tau) - \theta(t) - \theta'(t)(\tau - t)$ .

Direct calculations show that

$$(A.10) \quad \begin{aligned} \left( \frac{g(\tau, t)}{\theta'(\tau)} \right)' &= -\frac{i(\theta'(\tau) - \theta'(t))}{\theta'(\tau)} e^{i\Delta\theta(\tau, t)} \psi \left( \frac{\tau - t}{\omega} \right) + \frac{1}{\theta'(\tau)} (1 - e^{i\Delta\theta(\tau, t)}) \frac{1}{\omega} \psi' \left( \frac{\tau - t}{\omega} \right) \\ &\quad - \frac{g(\tau, t)\theta''(\tau)}{(\theta'(\tau))^2}. \end{aligned}$$

Among the three terms, the third one is easiest to bound. We have

$$(A.11) \quad \left| \frac{g(\tau, t)\theta''(\tau)}{(\theta'(\tau))^2} \right| \leq 2\epsilon \left| \psi \left( \frac{\tau - t}{\omega} \right) \right|.$$

It follows that

$$(A.12) \quad \left| \int_{\mathbb{R}} \frac{g(\tau, t)\theta''(\tau)}{(\theta'(\tau))^2} e^{-i\theta(\tau)} d\tau \right| \leq 2\epsilon \int_{\mathbb{R}} |\psi(\tau)| d\tau = 2I_1\epsilon.$$

To bound the other two terms, we need some preparations.

First, by using the assumption  $|(\frac{1}{\theta'})'| = |\frac{\theta''}{(\theta')^2}| \leq \epsilon$ , we can bound  $|\theta'(t) - \theta'(\tau)|$  by integrating the above inequality from  $t$  to  $\tau$  as follows:

$$(A.13) \quad \left| \frac{1}{\theta'(\tau)} - \frac{1}{\theta'(t)} \right| \leq \epsilon |\tau - t|,$$

which gives that

$$(A.14) \quad |\theta'(\tau) - \theta'(t)| \leq \epsilon \theta'(\tau) \theta'(t) |\tau - t| \quad \forall t, \tau \in \mathbb{R}.$$

We now estimate  $|1 - e^{i\Delta\theta(\tau,t)}|$  in two different ways:

$$(A.15) \quad \begin{aligned} |e^{-i\Delta\theta(\tau,t)} - 1| &\leq |\Delta\theta(\tau,t)| \leq |\theta'(t^*) - \theta'(t)||t - \tau| \leq \epsilon\theta'(t)\theta'(t^*)|t - \tau|^2 \\ &\leq M'\epsilon\theta'(\tau)\theta'(t)|t - \tau|^2, \end{aligned}$$

where  $t^*$  is a number between  $t$  and  $\tau$ , and also

$$(A.16) \quad \begin{aligned} |e^{-i\Delta\theta(\tau,t)} - 1| &\leq |\Delta\theta(\tau,t)| \leq |\theta'(t^*) - \theta'(t)||t - \tau| \leq \left( \sup_{t \in \mathbb{R}} \theta'(t) - \inf_{t \in \mathbb{R}} \theta'(t) \right) |t - \tau| \\ &\leq (M' - 1) \left( \inf_{t \in \mathbb{R}} \theta'(t) \right) |t - \tau| \leq (M' - 1)\theta'(t)|t - \tau|. \end{aligned}$$

The first term can be estimated as follows:

$$(A.17) \quad \begin{aligned} &\left| \int_{\mathbb{R}} \frac{i(\theta'(\tau) - \theta'(t))}{\theta'(\tau)} e^{i\Delta\theta(\tau,t)} \psi\left(\frac{\tau-t}{\omega}\right) e^{-i\theta(\tau)} d\tau \right| \\ &= \left| \int_{\mathbb{R}} \left(1 - \frac{\theta'(t)}{\theta'(\tau)}\right) \psi\left(\frac{\tau-t}{\omega}\right) e^{i\theta'(t)(\tau-t)} d\tau \right| \\ &= \left| \frac{1}{\theta'(t)} \int_{\mathbb{R}} \left[ \left(1 - \frac{\theta'(t)}{\theta'(\tau)}\right) \psi\left(\frac{\tau-t}{\omega}\right) \right]' e^{i\theta'(t)(\tau-t)} d\tau \right| \\ &\leq \left| \int_{\mathbb{R}} \frac{\theta''(\tau)}{(\theta'(\tau))^2} \psi\left(\frac{\tau-t}{\omega}\right) e^{i\theta'(t)(\tau-t)} d\tau \right| + \left| \frac{1}{\omega} \int_{\mathbb{R}} \frac{\theta'(\tau) - \theta'(t)}{\theta'(t)\theta'(\tau)} \psi'\left(\frac{\tau-t}{\omega}\right) e^{i\theta'(t)(\tau-t)} d\tau \right| \\ &\leq \epsilon|\omega| \left( \int_{\mathbb{R}} |\psi(\tau)| d\tau + \int_{\mathbb{R}} |\tau\psi'(\tau)| d\tau \right) \\ &\leq |\omega|(I_1 + I_2)\epsilon. \end{aligned}$$

The second equality is obtained by integration by parts. To get the second inequality, we use the assumptions that  $|\frac{\theta''(\tau)}{(\theta'(\tau))^2}| \leq \epsilon$  and (A.14).

Now, we estimate the second term. Let  $g_2(\tau, t) = \frac{1}{\theta'(\tau)}(1 - e^{i\Delta\theta(\tau,t)})\frac{1}{\omega}\psi'\left(\frac{\tau-t}{\omega}\right)$ . First, using the fact that  $\Delta\theta(\tau, t) = \theta(\tau) - \theta(t) - \theta'(t)(\tau - t)$ , we have

$$(A.18) \quad \begin{aligned} \left| \int_{\mathbb{R}} g_2(\tau, t) e^{-i\theta(\tau)} d\tau \right| &= \left| e^{-i\theta(t)} \int_{\mathbb{R}} \frac{e^{-i\Delta\theta(\tau,t)} - 1}{\omega\theta'(\tau)} \psi'\left(\frac{\tau-t}{\omega}\right) e^{-i\theta'(t)(\tau-t)} d\tau \right| \\ &= \left| \frac{1}{\theta'(t)} \int_{\mathbb{R}} \left[ \frac{e^{-i\Delta\theta(\tau,t)} - 1}{\omega\theta'(\tau)} \psi'\left(\frac{\tau-t}{\omega}\right) \right]' e^{-i\theta'(t)(\tau-t)} d\tau \right| \end{aligned}$$

and

$$(A.19) \quad \begin{aligned} &\frac{1}{\theta'(t)} \left[ \frac{e^{-i\Delta\theta(\tau,t)} - 1}{\omega\theta'(\tau)} \psi'\left(\frac{\tau-t}{\omega}\right) \right]' \\ &= -\frac{\theta''(\tau)}{(\theta'(\tau))^2} \frac{(e^{-i\Delta\theta(\tau,t)} - 1)}{\omega\theta'(t)} \psi'\left(\frac{\tau-t}{\omega}\right) - \frac{i(\theta'(\tau) - \theta'(t))}{\omega\theta'(\tau)\theta'(t)} e^{-i\Delta\theta(\tau,t)} \psi'\left(\frac{\tau-t}{\omega}\right) \\ &\quad + \frac{e^{-i\Delta\theta(\tau,t)} - 1}{\omega^2\theta'(\tau)\theta'(t)} \psi''\left(\frac{\tau-t}{\omega}\right). \end{aligned}$$

Then, (A.18) can be bounded term by term as follows:

(A.20)

$$\text{using (A.16), } \left| \frac{\theta''(\tau)}{(\theta'(\tau))^2} \frac{(e^{-i\Delta\theta(\tau,t)} - 1)}{\omega\theta'(t)} \psi' \left( \frac{\tau-t}{\omega} \right) \right| \leq (M' - 1)\epsilon \left| \frac{\tau-t}{\omega} \psi' \left( \frac{\tau-t}{\omega} \right) \right|,$$

(A.21)

$$\text{using (A.14), } \left| \frac{i(\theta'(\tau) - \theta'(t))}{\omega\theta'(\tau)\theta'(t)} e^{-i\Delta\theta(\tau,t)} \psi' \left( \frac{\tau-t}{\omega} \right) \right| \leq \epsilon \left| \frac{\tau-t}{\omega} \psi' \left( \frac{\tau-t}{\omega} \right) \right|,$$

(A.22)

$$\text{using (A.15), } \left| \frac{e^{-i\Delta\theta(\tau,t)} - 1}{\omega^2\theta'(\tau)\theta'(t)} \psi'' \left( \frac{\tau-t}{\omega} \right) \right| \leq M'\epsilon \left| \left( \frac{\tau-t}{\omega} \right)^2 \psi'' \left( \frac{\tau-t}{\omega} \right) \right|.$$

By combining these inequalities—(A.20), (A.21), and (A.22)—we get

$$(A.23) \quad \left| \int_{\mathbb{R}} g_2(\tau, t) e^{-i\theta(\tau)} d\tau \right| \leq M'|\omega|(I_2 + I_3)\epsilon,$$

where

$$(A.24) \quad I_3 = \int_{\mathbb{R}} |\tau^2 \psi''(\tau)| d\tau.$$

Then the proof is completed by combining (A.7), (A.12), (A.17), and (A.23).  $\square$

**Appendix B. Proof of Lemma 3.2.** To prove Lemma 3.2, we need the following technical lemma.

LEMMA B.1. *Let  $\epsilon \in (0, 1)$ , and let  $g(t)$  be a positive, continuous, and piecewise  $C^1$  function on  $[c, c + 2n\pi]$ , where  $n$  is an integer. Suppose*

$$(B.1) \quad \left| \frac{g'(t)}{g(t)} \right| < \epsilon \quad \forall t \in [c, d].$$

Then we have

$$(B.2) \quad \left| \int_c^{c+2n\pi} g(t) \cos t dt \right| < 2\pi\epsilon \int_c^{c+2n\pi} g(t) dt.$$

*Proof.* For each  $m \in \{0, 1, 2, \dots, n\}$ , let  $t_m = c + 2m\pi$ . We have

$$\begin{aligned} \left| \int_c^{c+2n\pi} g(t) \cos t dt \right| &= \left| \sum_{m=1}^n \int_{t_{m-1}}^{t_m} g(t) \cos t dt \right| = \left| \sum_{m=1}^n \int_{t_{m-1}}^{t_m} [g(t) - g(t_{m-1})] \cos t dt \right| \\ &= \left| \sum_{m=1}^n \int_{t_{m-1}}^{t_m} \left( \int_{t_{m-1}}^t g'(s) ds \right) \cos t dt \right| \\ &\leq \sum_{m=1}^n \int_{t_{m-1}}^{t_m} |\cos t| \left( \int_{t_{m-1}}^t |g'(s)| ds \right) dt \\ &\leq \sum_{m=1}^n \left( \int_{t_{m-1}}^{t_m} |\cos t| dt \right) \left( \int_{t_{m-1}}^{t_m} \epsilon g(s) ds \right) \\ &= \sum_{m=1}^n 4\epsilon \left( \int_{t_{m-1}}^{t_m} g(s) ds \right) \\ &= 4\epsilon \int_c^{c+2n\pi} g(s) ds. \end{aligned}$$

We complete the proof.  $\square$

Now, we can give the proof of Lemma 3.2.

*Proof of Lemma 3.2.* First, using  $\cos^2(\theta) = (1 + \cos(2\theta))/2$ , we get

$$(B.3) \quad \|a(t) \cos \theta(t)\|_{L^2}^2 = \frac{1}{2} \|a(t)\|_{L^2}^2 + \frac{1}{2} \int_0^1 a^2(t) \cos 2\theta(t) dt.$$

Let  $s = 2\theta(t)$ . Then we obtain

$$(B.4) \quad \int_0^1 a^2(t) \cos 2\theta(t) dt = \frac{1}{2} \int_{2\theta(0)}^{2\theta(1)} g(s) \cos s ds,$$

where  $t(s) := \theta^{-1}(\frac{s}{2})$  and

$$(B.5) \quad g(s) := \frac{a^2(t(s))}{\theta'(t(s))}.$$

So the derivative of  $g$  is

$$(B.6) \quad g'(s) = \frac{a(t(s))a'(t(s))}{[\theta'(t(s))]^2} - \frac{a^2(t(s)) \cdot \theta''(t(s))}{2[\theta'(t(s))]^3}.$$

Hence we get

$$(B.7) \quad \left| \frac{g'(s)}{g(s)} \right| = \left| \frac{a'(t(s))}{a(t(s)) \cdot \theta'(t(s))} - \frac{\theta''(t(s))}{2[\theta'(t(s))]^2} \right| < \frac{3}{2} \epsilon \quad \forall s \in [2\theta(0), 2\theta(1)].$$

Using Lemma B.1, we obtain

$$(B.8) \quad \left| \int_0^1 a^2(t) \cos 2\theta(t) dt \right| < 3\epsilon \int_{2\theta(0)}^{2\theta(1)} g(s) ds = 6\epsilon \int_0^1 a^2(t) dt.$$

The above estimate and (B.3) imply (3.16).

To prove (3.18), we represent the inner product in this inequality as follows:

$$(B.9) \quad \langle a \cos \theta, \bar{a} \cos \bar{\theta} \rangle = \frac{1}{2} \left[ \int_0^1 a(t) \bar{a}(t) \cos(\bar{\theta}(t) + \theta(t)) dt + \int_0^1 a(t) \bar{a}(t) \cos(\bar{\theta}(t) - \theta(t)) dt \right].$$

Let  $s = \theta(t) - \hat{\theta}(t)$ . We obtain

$$(B.10) \quad \int_0^1 a(t) \bar{a}(t) \cos(\bar{\theta}(t) - \theta(t)) dt = \int_{\bar{\theta}(0) - \theta(0)}^{\bar{\theta}(1) - \theta(1)} g(s) \cos s ds,$$

where  $t(s) = (\bar{\theta} - \theta)^{-1}(s)$  and

$$(B.11) \quad g(s) = \frac{a(t(s)) \bar{a}(t(s))}{\bar{\theta}'(t(s)) - \theta'(t(s))}.$$

Thus, we have

$$(B.12) \quad \frac{d}{ds} g(s) = \frac{(\bar{\theta}' - \theta')(a\bar{a}' + a'\bar{a}) - a\bar{a}(\bar{\theta}'' - \theta'')}{(\bar{\theta}'(t(s)) - \theta'(t(s)))^3}$$

and

$$(B.13) \quad \left| \frac{g'(s)}{g(s)} \right| < \frac{2\epsilon}{(1 - \beta^{-1})^2} \quad \forall s \in [\bar{\theta}(0) - \theta(0), \bar{\theta}(1) - \theta(1)].$$

Using Lemma B.1 again, we get

$$(B.14) \quad \left| \int_0^1 a(t) \bar{a}(t) \cos(\bar{\theta}(t) - \theta(t)) dt \right| < \frac{8\epsilon}{(1 - \beta^{-1})^2} \int_0^1 a(t) \bar{a}(t) dt.$$

Similarly, we can show

$$(B.15) \quad \left| \int_0^1 a(t) \bar{a}(t) \cos(\bar{\theta}(t) + \theta(t)) dt \right| < \frac{8\epsilon(\beta^2 + \beta + 1)}{(1 + \beta)^2} \int_0^1 a(t) \bar{a}(t) dt < 8\epsilon \int_0^1 a(t) \bar{a}(t) dt.$$

Thus (3.18) follows by combining (B.9), (B.14), and (B.15).  $\square$

#### REFERENCES

- [1] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.
- [2] E. J. CANDÈS AND T. TAO, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.
- [3] S. CHEN, D. L. DONOHO, AND M. SAUNDERS, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33–61.
- [4] C. K. CHUI AND H. N. MHASKAR, *Signal decomposition and analysis via extraction of frequencies*, Appl. Comput. Harmon. Anal., to appear; available online from <http://www.sciencedirect.com/science/article/pii/S1063520315000044>.
- [5] I. DAUBECHIES, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 61, SIAM, Philadelphia, 1992.
- [6] I. DAUBECHIES, J. LU, AND H. WU, *Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool*, Appl. Comput. Harmon. Anal., 30 (2011), pp. 243–261.
- [7] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
- [8] K. DRAGOMIRETSKIY AND D. ZOZZO, *Variational mode decomposition*, IEEE Trans. Signal Process., 62 (2014), pp. 531–544.
- [9] P. FLANDRIN, *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego, 1999.
- [10] J. GILLES, *Empirical wavelet transform*, IEEE Trans. Signal Process., 61 (2013), pp. 3999–4010.
- [11] T. Y. HOU AND Z. SHI, *Data-driven time-frequency analysis*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 284–308.
- [12] T. Y. HOU AND Z. SHI, *Sparse Time-Frequency Decomposition by Dictionary Learning*, preprint, arXiv:1311.1163v2 [cs.IT], 2014.
- [13] T. Y. HOU, Z. SHI, AND P. TAVALLALI, *Convergence of a data-driven time-frequency analysis method*, Appl. Comput. Harmon. Anal., 37 (2014), pp. 235–270.
- [14] N. E. HUANG, Z. SHEN, S. R. LONG, M. C. WU, H. H. SHIH, Q. ZHENG, N.-C. YEN, C. C. TUNG, AND H. H. LIU, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 454 (1998), pp. 903–995.
- [15] S. MALLAT, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed., Elsevier/Academic Press, Amsterdam, 2009.
- [16] S. G. MALLAT AND Z. ZHANG, *Matching pursuits with time-frequency dictionaries*, IEEE Trans. Signal Process., 41 (1993), pp. 3397–3415.
- [17] Z. WU AND N. E. HUANG, *Ensemble empirical mode decomposition: A noise assisted data analysis method*, Adv. Adapt. Data Anal., 1 (2008), pp. 1–41.